Maximum likelihood estimator for multinomial model

$E=\{1,2,\dots,r\}$. pmf $\underline{p_j}$, $j=1,\dots,r$. Observe $X_1,\dots,X_n$ iid. $\mathbb{P}(X_i=j)=p_j$, $i=1,\dots,n$, $j=1,\dots,r$

Goal: ① Compute max. lh. estimator $\hat{p}$ for p

② Show that $\hat{p}$ asymptotically normal, find asymptotic covariance matrix

Parameter space: $\left\{p : p_j \geq 0,\ \sum_{j=1}^{r} p_j = 1\right\} = \mathcal{P}$

① $$\mathbb{P}(X_1=x_1,\dots,X_n=x_n)=\prod_{i=1}^{n}\mathbb{P}(X_i=x_i)=\prod_{i=1}^{n}\prod_{j=1}^{r}p_j^{\mathbb{1}\{x_i=j\}}=\prod_{j=1}^{r}\prod_{i=1}^{n}p_j^{\mathbb{1}\{x_i=j\}}$$

$$=\prod_{j=1}^{r}p_j^{\overbrace{\sum_{i=1}^{n}\mathbb{1}\{x_i=j\}}^{T_j}}=\prod_{j=1}^{r}p_j^{T_j}$$

$$\log lh=\log\mathbb{P}(X_1=x_1,\dots,X_n=x_n)=\log\prod_{j=1}^{r}p_j^{T_j}=\sum_{j=1}^{r}T_j\log p_j$$

Maximum likelihood estimator for multinomial model

$$\log \mathrm{lh} = \underbrace{\sum_{j=1}^{r} T_j \log p_j}_{=f(p)}, \quad T_j = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\}, \quad \mathcal{P} = \{p \in \mathbb{R}^r : p_j \geq 0 \ \forall j, \ \sum_{j=1}^{r} p_j = 1\}$$

① Calculating MLE: $\max\limits_{p \in \mathcal{P}} f(p) \Leftrightarrow \max f(p)$ s.t. $h(p) = \sum_{j=1}^{r} p_j - 1 = 0$

Assume $T_j > 0 \ \forall j$

Necessary conditions: $0 = \nabla f(\hat{p}) + \lambda \cdot \nabla h(\hat{p}), \quad \lambda \in \mathbb{R}$

$$\partial_{p_j} f(p) = \frac{T_j}{p_j} \overset{!}{=} 0, \quad p_j \to +\infty \ ???$$

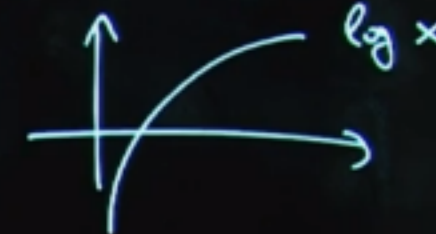
$$\partial_{p_j} h(p) = 1 \Rightarrow 0 = \frac{T_j}{\hat{p}_j} + \lambda \Rightarrow \lambda \neq 0 \Rightarrow \hat{p}_j = -\frac{T_j}{\lambda}$$

$$1 = \sum_{j=1}^{r} \hat{p}_j = \sum_{j=1}^{r} \left(-\frac{T_j}{\lambda}\right) = -\frac{1}{\lambda} \sum_{j=1}^{r} T_j = -\frac{n}{\lambda} \Rightarrow \lambda = -n \Rightarrow \boxed{\hat{p}_j = \frac{T_j}{n}}$$

Maximum likelihood estimator for multinomial model

$f(p) = \log \mathcal{L}_n = \sum_{j=1}^{r} T_j \log p_j$, $T_j = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\}$, $\mathcal{P} = \{ p \in \mathbb{R}^r : p_j \geq 0 \ \forall j, \sum_{j=1}^{r} p_j = 1 \}$ (simplex)

$h(p) = \sum_{j=1}^{r} p_j - 1$, $\partial_{p_j} f(p) = \frac{T_j}{p_j}$, $\partial_{p_j} h(p) = 1$



(1) $\underline{T_j > 0} \Rightarrow \hat{p}_j = \frac{T_j}{n}$. Global max?

$$\partial_{p_k} \partial_{p_j} f(p) = \partial_{p_k} \frac{T_j}{p_j} = \begin{cases} -\frac{T_j}{p_j^2}, & j = k \\ 0 & j \neq k \end{cases} \Rightarrow \nabla^2 f(p) < 0 \Rightarrow f \text{ concave}$$

$\underline{T_j = 0}$? (Karush-Kuhn-Tucker conditions)

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{j=1}^{r} p_j^{T_j} \Rightarrow \boxed{\hat{p}_j = \frac{T_j}{n}}$$ global maximum.

Maximum likelihood estimator for multinomial model

$\boxed{\hat{p}_j = \frac{T_j}{n}}$, $\mathbb{P}(X_i = j) = p_j$, $\mathcal{P} = \{p : p_j \geq 0, \sum p_j = 1\}$, $T_j = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\}$

(2) $\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{1}\{X_i = j\}}_{=(Y_i)_j}$, $\mathbb{E}[(Y_1)_j] = \mathbb{E}[\underbrace{\mathbb{1}\{X_1 = j\}}_{\sim Be(p_j)}] = \mathbb{P}(X_1 = j) = p_j$

CLT: $\sqrt{n} \cdot (\hat{p} - \mathbb{E}[Y_1]) \xrightarrow[n \to \infty]{D} \mathcal{N}(0, \underbrace{Cov(Y_1)}_{= \Sigma})$

$Var((Y_1)_j) = p_j(1 - p_j)$

$\mathbb{E}[(Y_1)_j (Y_1)_k] = \mathbb{E}[\mathbb{1}\{X_1 = j\} \mathbb{1}\{X_1 = k\}] = 0, \quad j \neq k$

$Cov((Y_1)_j, (Y_1)_k) = \mathbb{E}[(Y_1)_j (Y_1)_k] - \mathbb{E}[(Y_1)_j] \mathbb{E}[(Y_1)_k]$

$= 0 - p_j \cdot p_k, \quad j \neq k$

$$\Sigma_{j,k} = \begin{cases} Var((Y_1)_j) & , j = k \\ Cov((Y_1)_j, (Y_1)_k) & , j \neq k \end{cases}$$

$$= \begin{cases} p_j(1 - p_j) & , j = k \\ -p_j p_k & , j \neq k \end{cases}$$

Maximum likelihood estimator for multinomial model

$$\hat{p}_j = \frac{T_j}{n}, \quad T_j = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\}, \quad \mathcal{P} = \{p : p_j \geq 0, \sum_{j=1}^{r} p_j = 1\}$$

② $\hat{\Theta}$ MLE, log lh $f(\Theta)$

$$\sqrt{n}(\hat{\Theta} - \Theta^*) \xrightarrow[n \to \infty]{D} N(0, I(\Theta^*)^{-1})$$

$$I(\Theta) = -\mathbb{E}[\nabla^2 f(\Theta)]$$

$$f(p) = \sum_{j=1}^{r} T_j \cdot \log p_j, \quad \partial_{p_j} f(p) = \frac{T_j}{p_j}, \quad \boxed{n=1}$$

$$\partial_{p_j}\partial_{p_k} f(p) = \begin{cases} -\frac{T_j}{p_j^2}, & j = k \\ 0, & j \neq k \end{cases}$$

$$I(p)_{jk} = -\mathbb{E}[(\nabla^2 f(p))_{jk}] = \begin{cases} \frac{p_j}{p_j^2} = \frac{1}{p_j}, & j = k \\ 0, & j \neq k \end{cases}$$

$$I(p)^{-1}_{jk} = \begin{cases} p_j, & j = k \\ 0, & j \neq k \end{cases}$$

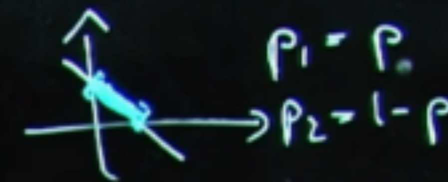
$\neq$

$$\Sigma_{j,k} = \begin{cases} p_j(1 - p_j), & j = k \\ -p_j p_k, & j \neq k \end{cases}$$

???

Maximum likelihood estimator for multinomial model

$\hat{p}_j = \frac{T_j}{n}$, $T_j = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\}$, $\tilde{\mathcal{P}} = \{p : p_j > 0, \boxed{\sum_{j=1}^{r} p_j = 1}\} \subseteq \mathbb{R}^r$



2' Thm: $\hat{\Theta}$ MLE, $\log$ lh $f(\Theta)$

1. Model identified ✓
2. supp $\mathbb{P}_\Theta$ does depend on Θ ✓
3. Θ^* does not lie on the boundary of Θ ✗
4. $I(\Theta)$ is invertible in a neighborhood of Θ^* ✗
5. Technical assmpts.

$$\sqrt{n}(\hat{\Theta} - \Theta^*) \xrightarrow[n \to \infty]{D} N(0, I(\Theta^*)^{-1})$$

$$I(\Theta) = -\mathbb{E}[\nabla^2 f(\Theta)]$$

$$\Sigma \neq I(p)^{-1}$$

$$\tilde{\tilde{\mathcal{P}}} = \{p \in \mathbb{R}^{r-1} : p_j > 0, \underbrace{1 - \sum_{j=1}^{r-1} p_j}_{p_r} > 0\}$$

$$\tilde{\tilde{I}}(p)^{-1} = (\Sigma)_{\substack{j=1,\dots,r-1 \\ k=1,\dots,r-1}}$$